

Towards a better understanding of Random Forest mechanisms

Simon Bernard (Lab. LITIS, Université de Rouen) 17/02/2009.

Random Forest (RF) is a family of classifier ensemble methods that use randomization to produce a diverse pool of individual classifiers, as for Bagging [Breiman96] or Random Subspaces methods [Ho98]. Those methods have shown to be particularly competitive with one of the most efficient learning principles, i.e. boosting [Breiman01,CZ01,RKA06]. However, the mechanisms that explain the good performance of this type of classifier ensemble are not clearly identified and one has to admit that it is still a complex task for the practitioner to take full benefits of the potential of those methods. In this talk we present some of our studies that aim at better understanding RF mechanisms and explaining why they perform well or not.

In the first part of my presentation I will deal with the parametrization of RF and more particularly the parametrization of the reference algorithm Forest-RI. I will present an experimental study on different parametrization solutions of this algorithm according to resulting RF performance.

In the second part, I will tackle the analysis of RF mechanisms through a classifier selection problem. By selecting different subsets of trees among a larger pool of trees, I will firstly show that better RF can be obtained even using a sub-optimal classifier selection method. Then I will present some of our experimental investigations to find an explanation of these differences of accuracy.

References:

- [Breiman96] L. Breiman, “Bagging predictors”, *Machine Learning* 24(2), pp. 123–140, 1996.
- [Breiman01] L. Breiman, “Random forests”, *Machine Learning* 45(1), pp. 5–32, 2001.
- [CZ01] A. Cutler and G. Zhao. “Pert - perfect random tree ensembles”, *Computing Science and Statistics* 33, 2001.
- [Ho98] T. Ho, “The random subspace method for constructing decision forests”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8), pp. 832–844, 1998.
- [Rodriguez et al., 2006] J.J. Rodriguez, L.I. Kuncheva, and C.J. Alonso,

“Rotation forest : A new classifier ensemble method”, IEEE Trans. Pattern Analysis and Machine Intelligence 28(10), pp. 1619–1630, 2006.